Winning Space Race with Data Science

Wentao Lu
2022.07.18

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Summary of methodologies
  - The project used Web Scraping and SpaceX API to gather necessary data
  - Data analysis including data wrangling, visualization and interactive visual dashboard
  - Machine Learning Algorithms using KNN, Decision Tree, SVM etc.

- Summary of all results
  - The result of data prediction is in overall 80 – 90% accuracy

Introduction

- Project background and context
  - The newly founded company Space Y is to compete with Space X
- The best way to wining the competition is to have lower cost of launches by analyzing Space X landings. With recyclable launches, it is the key point to be successfully

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Space X API
  - Wki web scraping
- Perform data wrangling
  - creating landing outcome label
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Decision Tree, KNN, SVM etc…

Data Collection

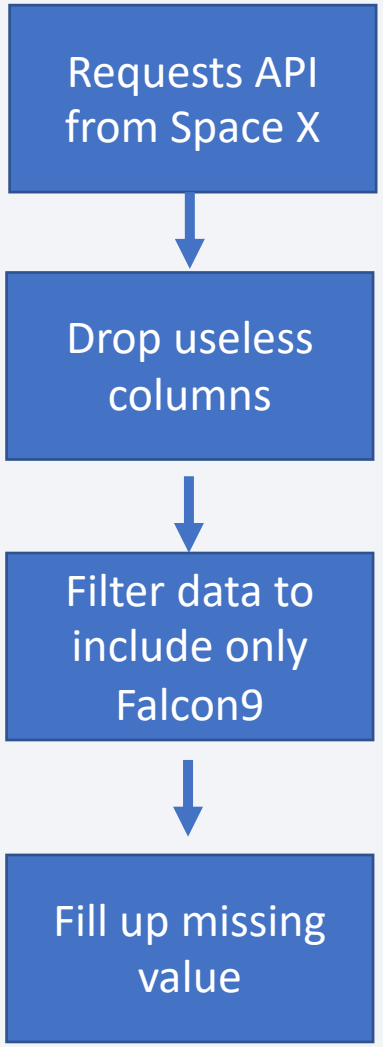
Datasets are collected via Space X API via https://api.spacexdata.com/v4

and Wikipedia via

Data Collection – SpaceX API

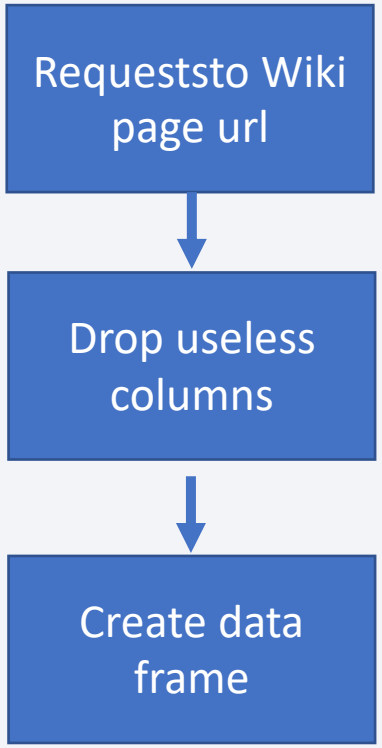
- Space X API

- Refer https://github.com/lwt0000/Applied-Data-Science-Capstone



Data Collection - Scraping

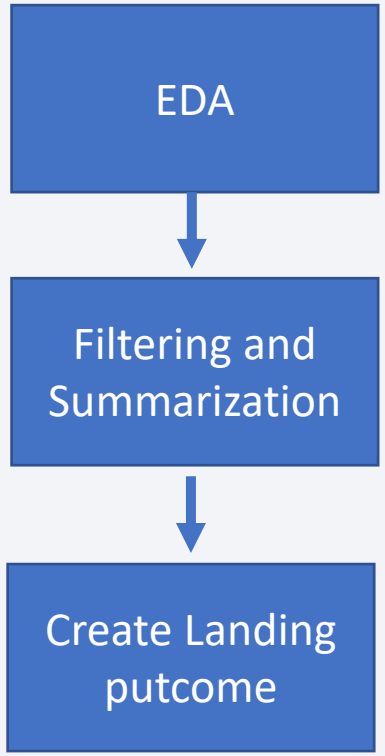
- Web Scraping
- Refer https://github.com/lwt0000/Applied-Data-Science-Capstone



Data Wrangling

First Using Exploratory Data Analysis EDA to analysis data

Then Filtering useful info and create labels



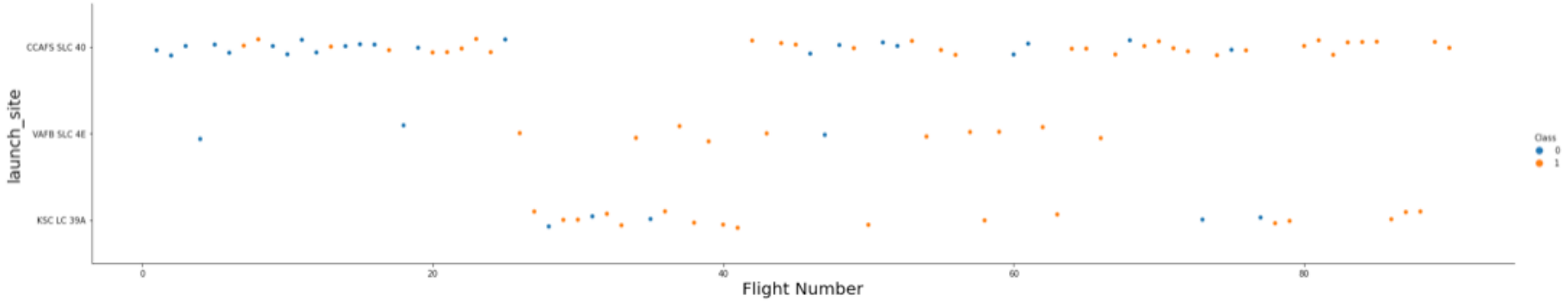
Refer https://github.com/lwt0000/Applied-Data-Science-Capstone

EDA with Data Visualization

TASK 1: Visualize the relationship between Flight Number and Launch Site

Use the function `catplot` to plot `FlightNumber` vs `LaunchSite`, set the parameter `x` parameter to `FlightNumber`,set the `y` to `Launch Site` and set the parameter `hue` to `'class'`

```
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("launch_site",fontsize=20)
plt.show()
```



Now try to explain the patterns you found in the Flight Number vs. Launch Site scatter point plots.

Refer https://github.com/lwt0000/Applied-Data-Science-Capstone

EDA with SQL

- **Display the names of the unique launch sites in the space mission**
- **Display 5 records where launch sites begin with the string 'CCA'**
- **Display the total payload mass carried by boosters launched by NASA (CRS)**
- **Display average payload mass carried by booster version F9 v1.1**
- **List the date when the first succesful landing outcome in ground pad was acheived.**
- **List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000**
- **List the total number of successful and failure mission outcomes**
- **List the names of the booster_versions which have carried the maximum payload mass. Use a subquery**
- **List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.**
- **Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.**

Refer https://github.com/lwt0000/Applied-Data-Science-Capstone

Build an Interactive Map with Folium

- Makers, circles, lines and clusters were used in the Folium map
  - Using maker to indicate the launch site
  - Circles indicate highlighted areas

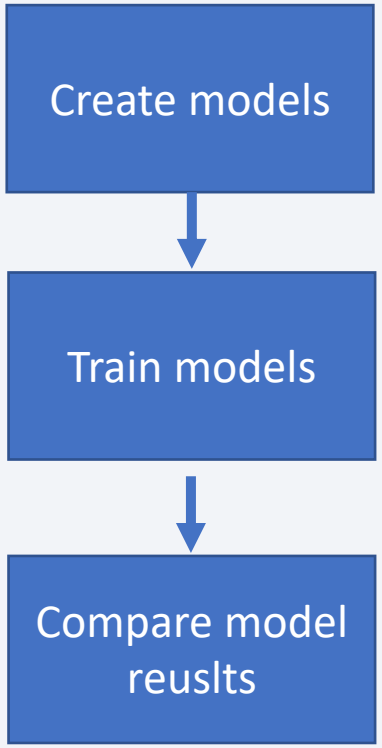
Refer https://github.com/lwt0000/Applied-Data-Science-Capstone

Build a Dashboard with Plotly Dash

- Percentage of launches by site and Payload range graph are used to plot the graph
- These allow us to quickly understand the data trends

Refer https://github.com/lwt0000/Applied-Data-Science-Capstone

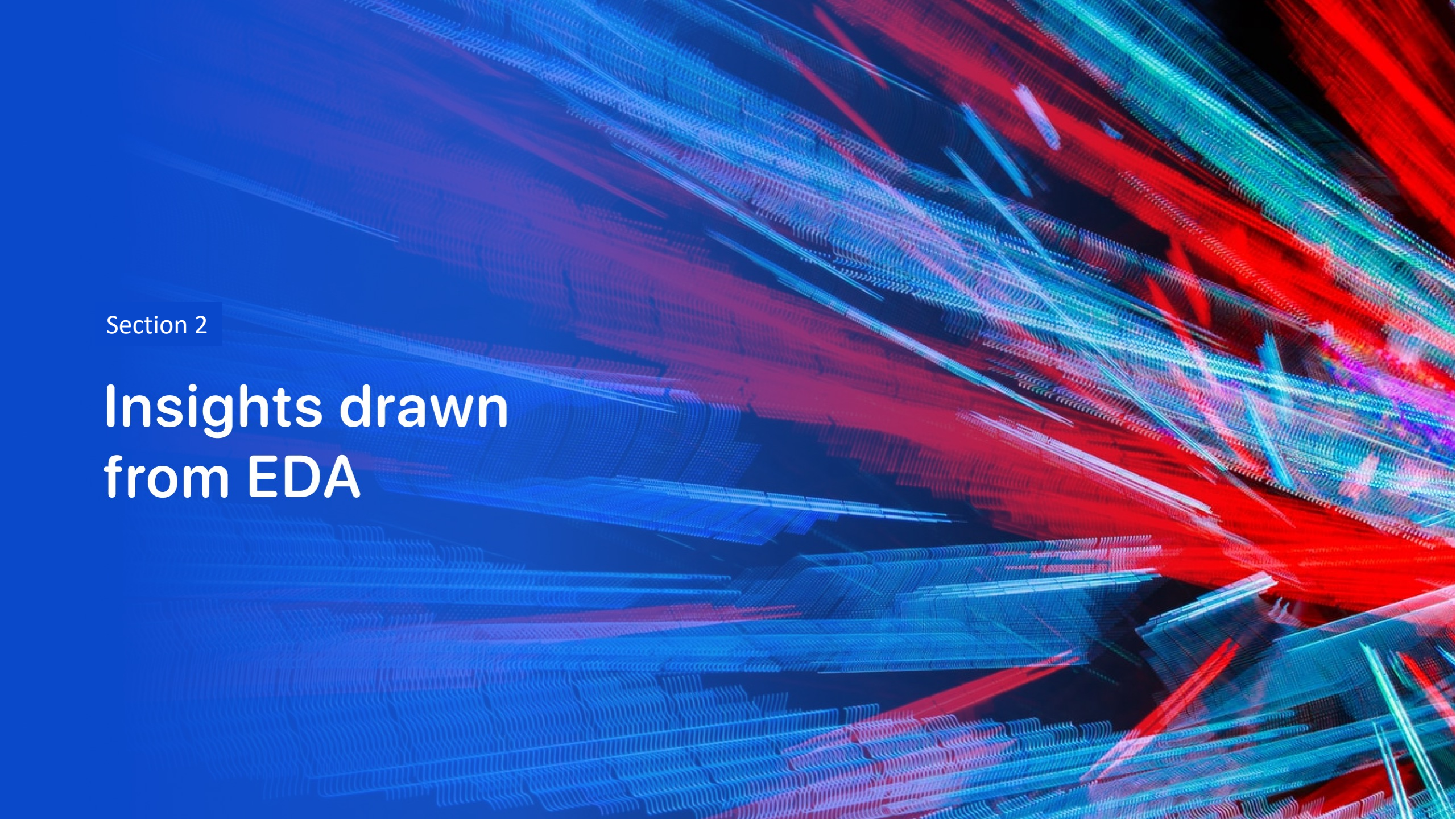
Predictive Analysis (Classification)



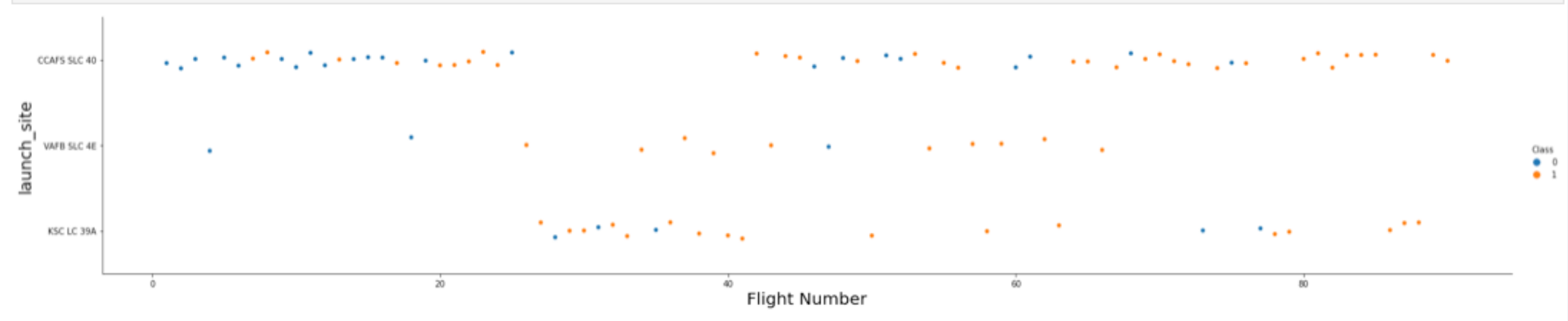
Refer https://github.com/lwt0000/Applied-Data-Science-Capstone

Results

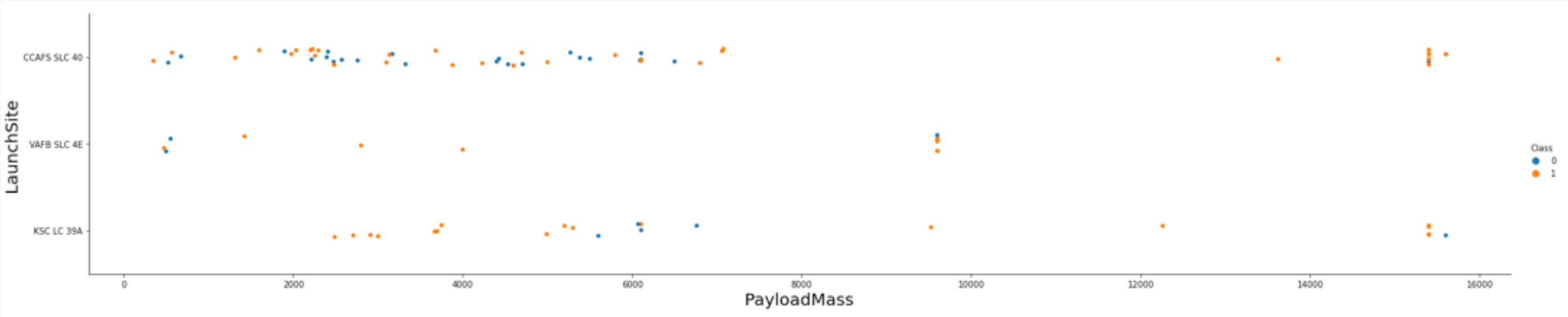
- Space X use 4 launch sites
- Landing can be on ocean and ground
- F9 v1.1 payload average is 2928kg


Section 2
Insights drawn from EDA

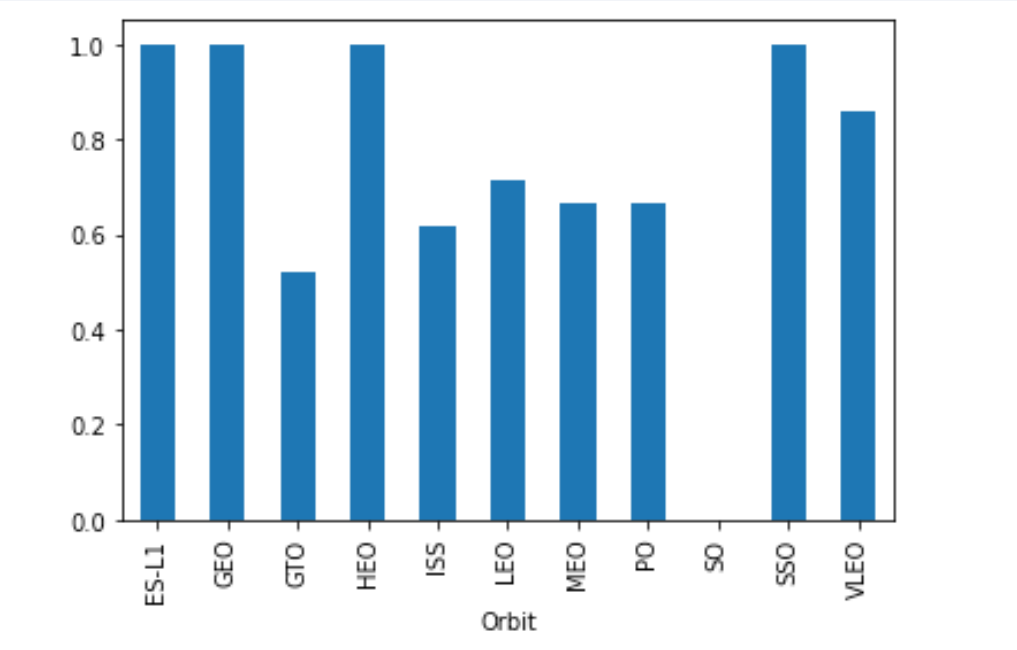
Flight Number vs. Launch Site



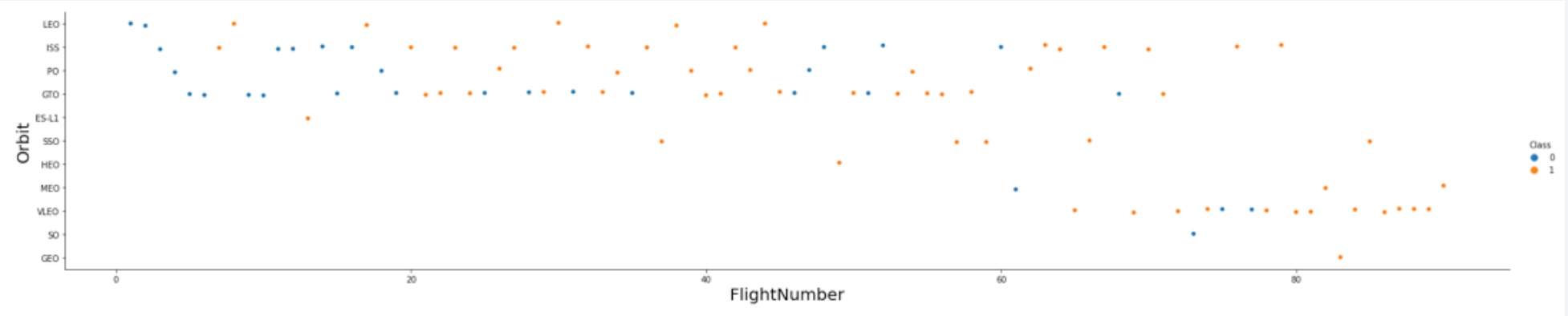
Payload vs. Launch Site



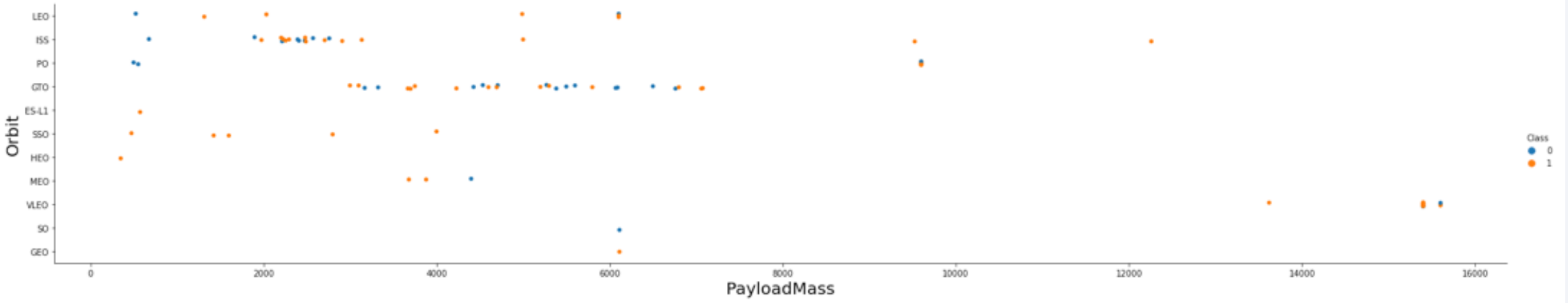
Success Rate vs. Orbit Type



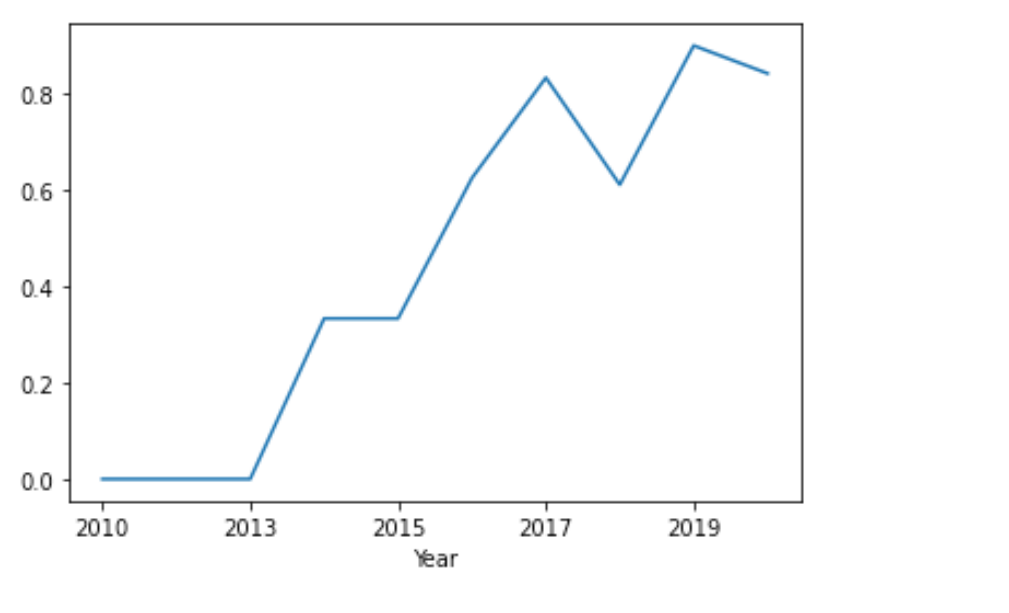
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

```
sql SELECT * FROM SPACEXTBL WHERE Launch_SITE LIKE 'CCA%' LIMIT 5;
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

```
sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD FROM SPACEXTBL WHERE PAYLOAD  LIKE '%CRS%';
```

```
 * sqlite:///my_data1.db
Done.
```

TOTAL_PAYLOAD
111268

Average Payload Mass by F9 v1.1

```
sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD FROM SPACEXTBL WHERE Booster_Version LIKE 'F9 V1.1';
```

```
 * sqlite:///my_data1.db
Done.
```

AVG_PAYLOAD
2928.4

First Successful Ground Landing Date

first_success_gp
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

booster_version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

mission_outcome	qty
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

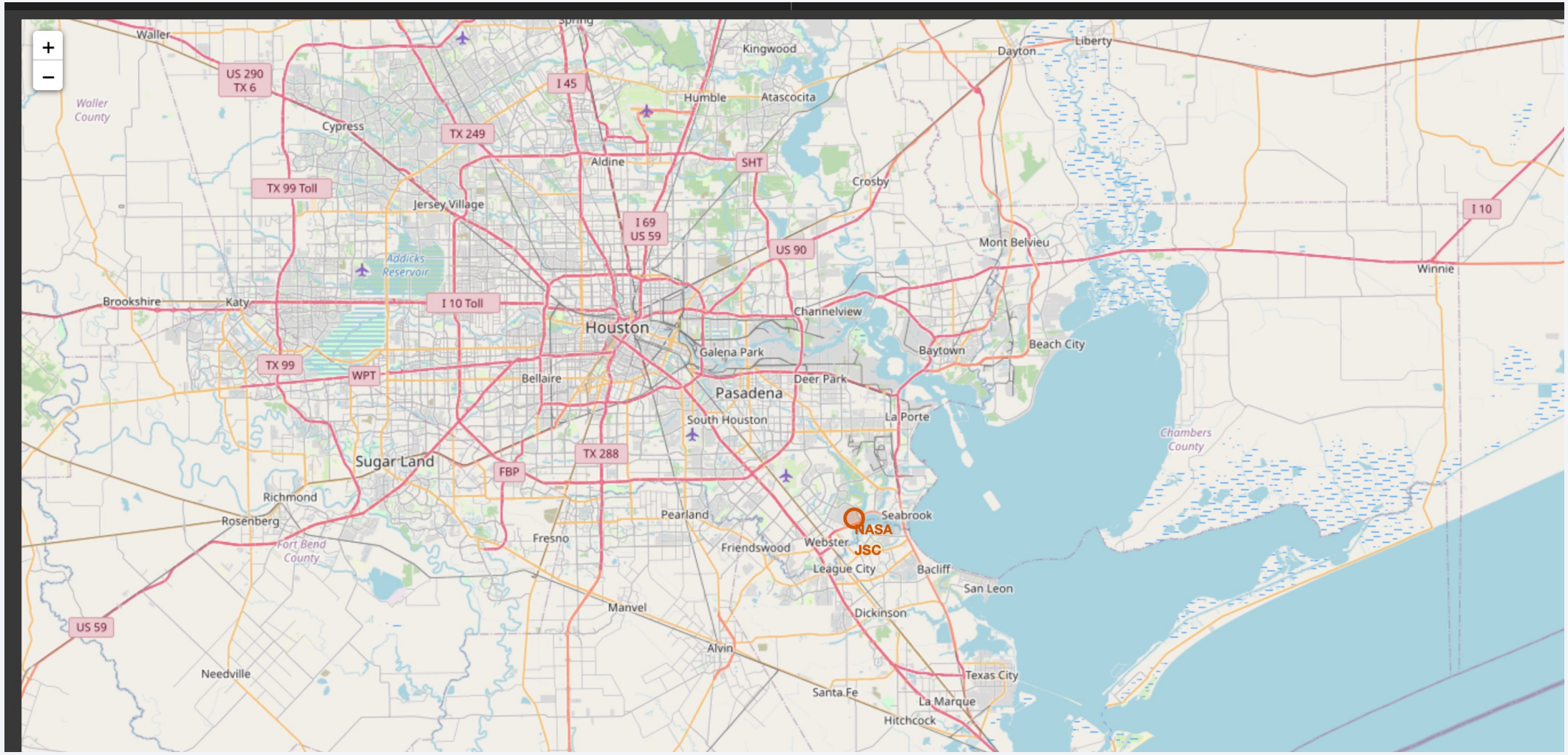
booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

landing__outcome	qty
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

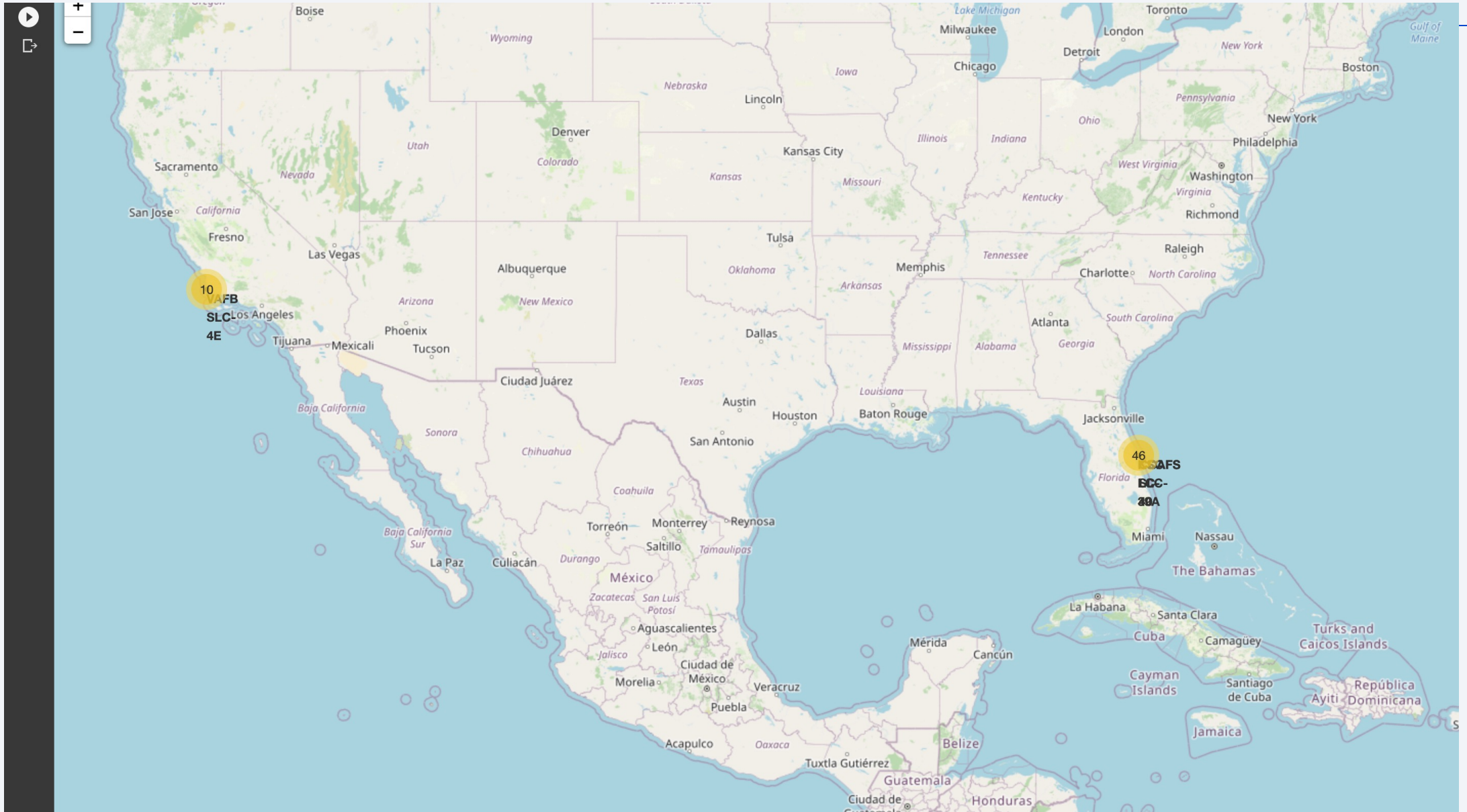

Section 3
Launch Sites
Proximities Analysis

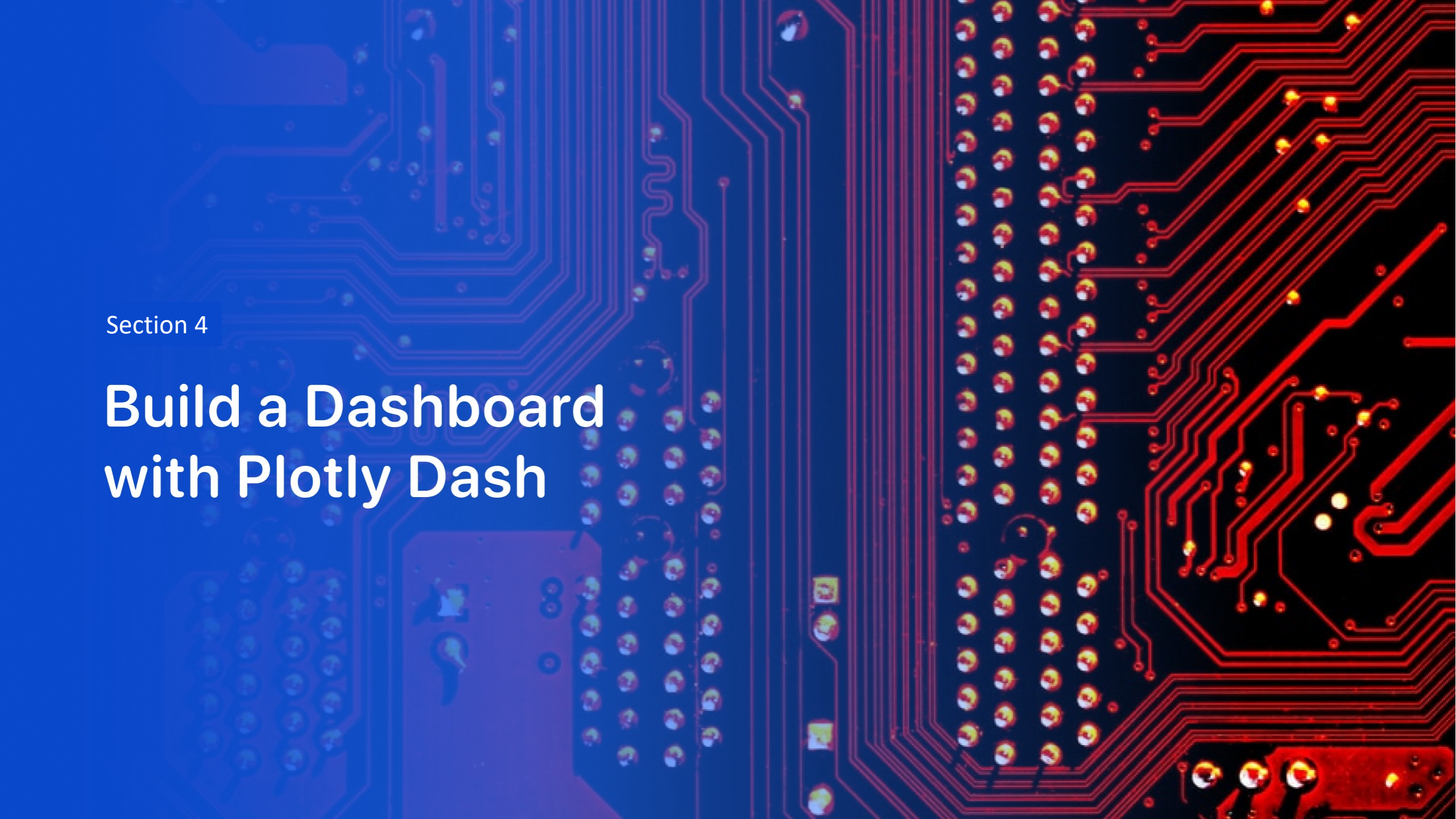
<Folium Map Screenshot 1>



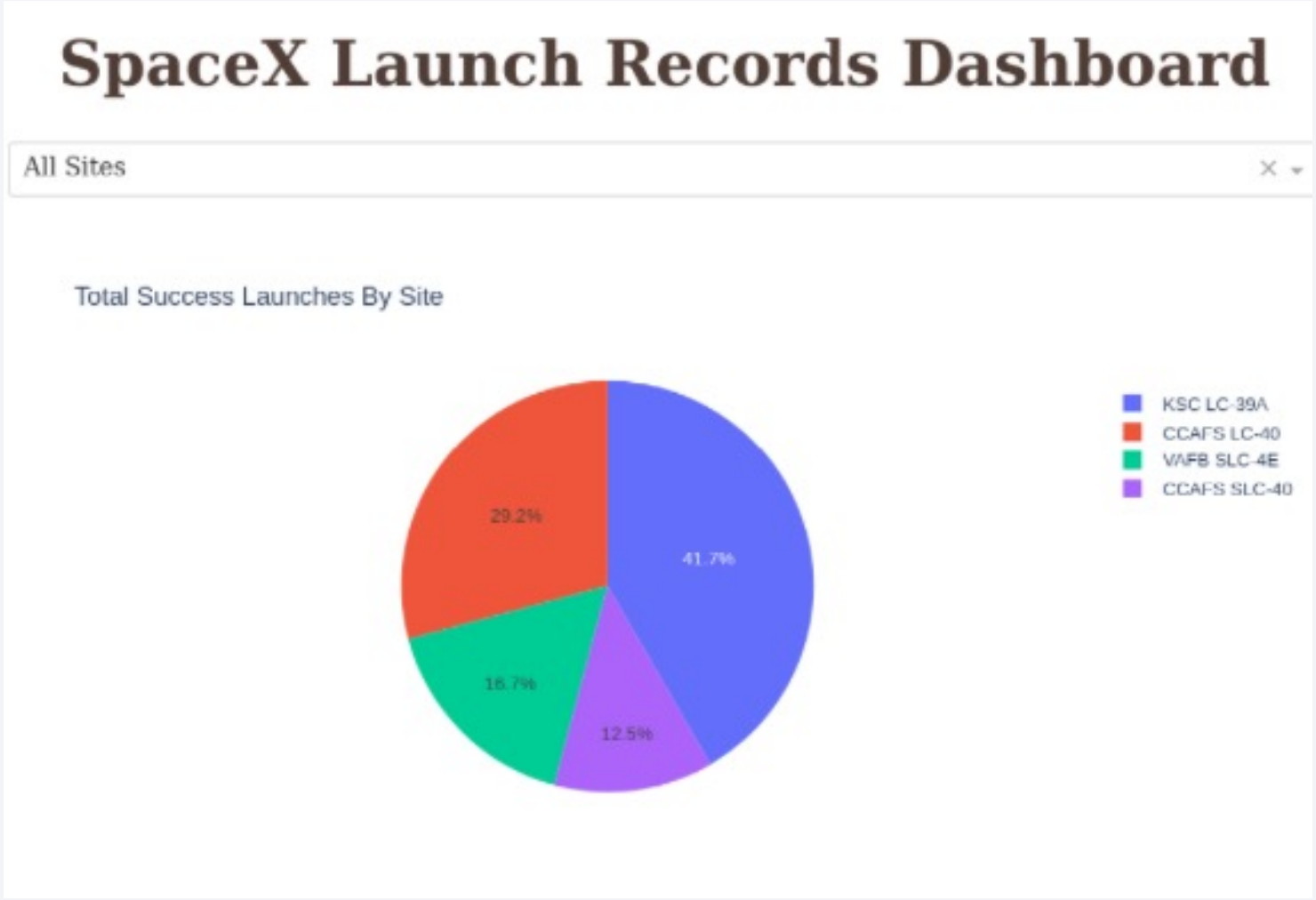
<Folium Map Screenshot 2>

<Folium Map Screenshot 3>

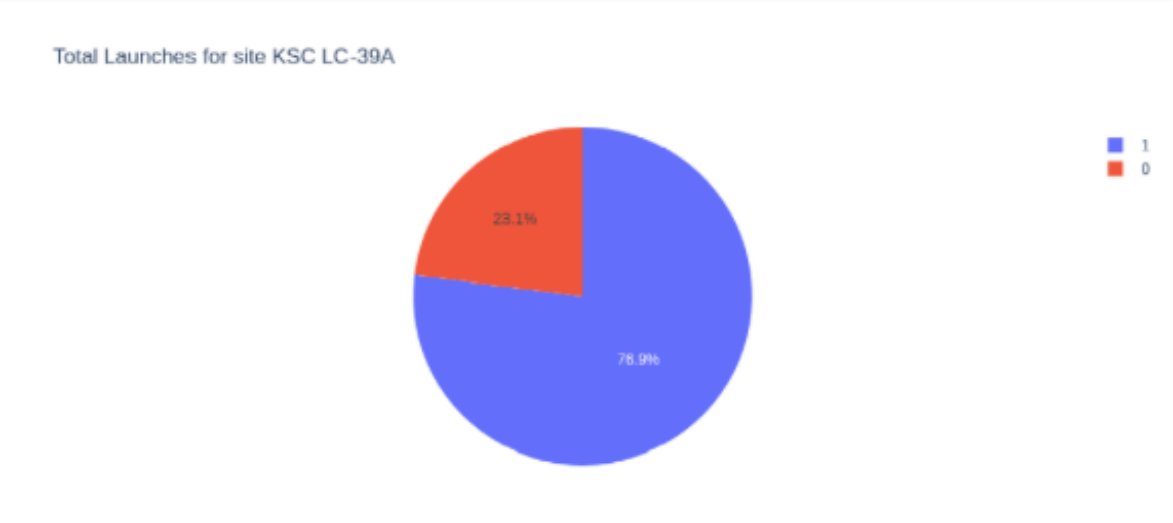



Section 4
Build a Dashboard with Plotly Dash

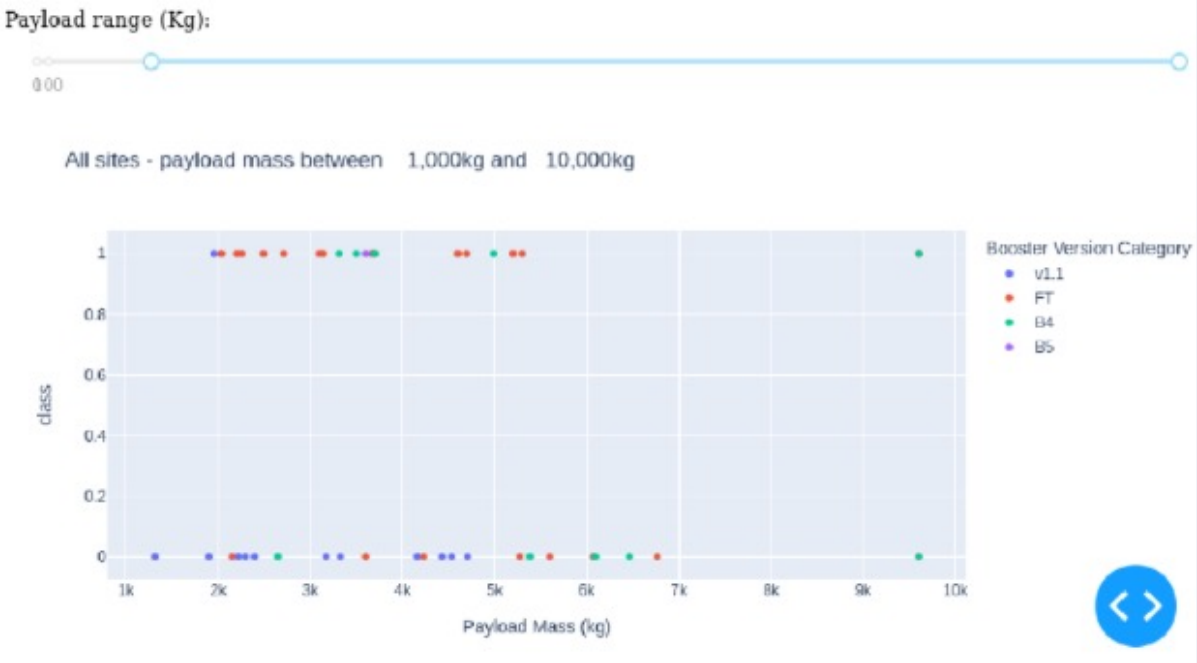
Successful Launches by Site



Launch Success Ratio For KSC LC 39A



Payload vs. Launch Outcome



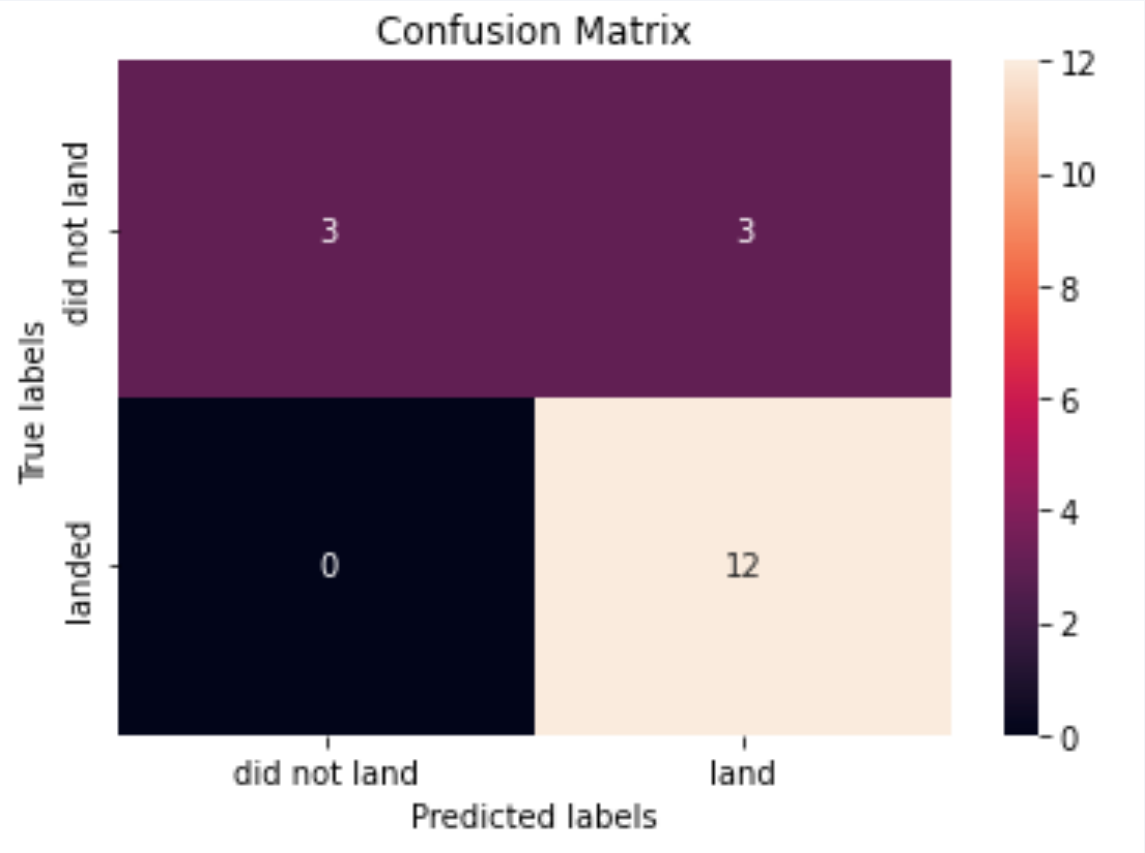
Section 5

Predictive Analysis (Classification)

Classification Accuracy

Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.88929	0.72222
KNN	0.84821	0.83333

Confusion Matrix



Conclusions

- The best launch site is KSC LC-39A;
- Most of mission outcomes are successful
- successful landing improves over time, Decision Tree Classifier can be used to predict successful landings

Appendix

Thank you!